Chapter 2

Evolved *Escherichia coli* strains for amplified, functional expression of membrane proteins

---

Nadia Gul[#], Daniel Linares-Martinez[#] and Bert Poolman.

# These authors contributed equally to the work.

1. Abstract

Membrane proteins play a pivotal role in a variety of essential cellular functions. The major barrier to the physical characterization and structure determination of membrane proteins is low protein yield and / or low functionality in recombinant expression. The enteric bacterium *E. coli* is the most widely employed organism for producing recombinant proteins. Beside several advantages of this expression host, one major drawback is that the protein of interest does not always adopt its native conformation and may end up in large insoluble aggregates. We describe a robust strategy to increase the likelihood of overexpressing membrane proteins in a functional state. The method involves fusion in tandem of GFP and the erythromycin-resistance protein (23S rRNA adenine N-6 methyl-transferase, ErmC) C-terminal of a target membrane protein. The fluorescence of GFP is used to report the folding state of the target protein, whereas ErmC is used to select for increased expression. By gradually increasing the erythromycin concentration of the medium and testing different membrane protein targets, we obtained a number of evolved strains of which four (NG2, NG3, NG5 and NG6) were characterized and their genome fully sequenced. The strains contained two or three mutations. Strikingly, each of the strains carried a mutation in the *hns* gene, whose product is involved in genome organization and transcriptional silencing. The degree of expression of (membrane) proteins correlates with the severity of the *hns* mutation. We propose that removal of the transcriptional silencing mechanism increases the levels of proteins essential for the functional overexpression of membrane proteins.

2. Introduction

Membrane proteins play a crucial role in a wide range of physiological processes including metabolic energy conservation, nutrient uptake, detoxification, regulation of cell volume and ion homeostasis, and signaling of environmental conditions. Membrane proteins represent 20-30% of the proteome of prokaryotic and eukaryotic organisms, and they comprise more than seventy percent of the current drug targets (Krogh et al. 2001, Khnouf et al. 2010, Lundstrom 2004). However, our understanding of function-structure relationships of membrane proteins is low as compared to water-soluble proteins. This imbalance has been attributed to the unparalleled difficulties of obtaining sufficient amounts of protein for functional and structural studies proteins. It has been noted that the most optimal host for overexpression is the native source (Grisshammer, Tate 1995, Bill et al. 2011). However, this is not a feasible strategy for most membrane proteins since they are either not expressed at a high enough level or a suitable self-expression technology is not available (Shaw, Miroux 2003). We thus need new and improved heterologous expression systems. For large-scale production of proteins, bacterial and fungal expression hosts such as *Escherichia coli*, *Lactococcus lactis, Saccharomyces cerevisiae* and *Pichia pastoris* are preferred (Steen et al. 2011, Oberg et al. 2009, Newstead et al. 2007).

Protein synthesis rates are generally much faster in prokaryotes than in eukaryotes (Widmann, Christen 2000), which can be advantageous but also cause problems in the downstream steps of membrane protein biogenesis (Bill et al. 2011). Although every organism has its pros and cons for the production of complex (multi-domain) membrane proteins, *E. coli* remains the most widely used host for the heterologous expression of proteins in high quantities (Mergulhao, Summers & Monteiro 2005, Raman, Cherezov & Caffrey 2006, Yin et al. 2007). The genetics and the fundamental understanding of transcription, translation and protein folding are far better characterized than for any other (micro)organism.

The challenge of overproducing membrane proteins in a functionally competent state is enormous. The available literature suggests that many membrane proteins can be expressed in reasonable amounts, but often they fail to get inserted into the membrane in their native conformations and are non-functional (Georgiou, Valax 1996, Klepsch, Persson & de Gier 2011). The difficulties to overproduce functional membrane proteins have been attributed to factors such as (i) toxicity of the proteins to the host cell (Laible et al. 2004); (ii) mismatch in codon usage and tRNA abundance (Zhang, Hubalewska & Ignatova 2009) (iii) limitation in the availability of precursors for protein synthesis (e.g. amino acids) (Marreddy et al. 2010); (iv) saturation of the membrane protein insertion machinery (Loll 2003, Wagner et al. 2006); (v) insufficient chaperone and foldase activity (Higgins, Demir & Tate 2003, Tate, Whiteley & Betenbaugh 1999); and (vi) limited membrane space to accommodate extra protein (Arechaga et al. 2000).

Compared soluble proteins, membrane proteins have a more complex biogenesis pathway and bottlenecks in their synthesis may arise from improper targeting, membrane insertion and/or folding (Georgiou, Valax 1996, Klepsch, Persson & de Gier 2011). The level of correctly-folded membrane protein can often be enhanced by systematic optimization of several parameters such as (i) directed evolution or engineering of the expression host (Wagner et al. 2008) (ii) tuning of expression conditions such as temperature, growth medium and inducer strength (Marreddy et al. 2010) (iii) modification of target proteins (Zoonens, Miroux 2010, Wang et al. 2003). Such screening is often done by trial-and-error and time-consuming. More efficient and target-specific screening methods are needed to tackle the enormous expression space (Schlegel et al. 2010). To date, one can simply screen for expression of membrane proteins in an overall well-folded state by fusion of a fluorescent reporter to the C-terminus of the target protein. If the target protein folds correctly the reporter is likely to reach its native conformation and become fluorescent (Waldo et al. 1999). A target protein that misfolds and aggregates will interfere with the folding of the reporter and be non-fluorescent. Thus, by monitoring the fluorescence of the reporter protein, one obtains a first indication of the folding state of the target protein. Earlier work has shown a very good correlation between fluorescence and the amount of membrane protein that is stably inserted into the membrane and functional (Drew et al. 2006, Drew et al. 2008, Geertsma et al. 2008a). We note that this correlation may not necessarily hold for every protein, i.e. in some cases a polypeptide may initiate folding and the

protein may insert into membrane but functionality may be compromised due to incorrect folding of sub-domains. In our hands, the GFP reporter strategy has worked well and forms the basis for the here-developed expression screening strategy.

In this study, we use an evolutionary strategy to optimize *E. coli* as host for the production of functional membrane protein. We combine the validated folding reporter GFP with the erythromycin-resistance protein (ErmC) to select for increased expression (figure 1A). With different target proteins, we find the generic transcriptional silencer HNS as factor that limits the overexpression of membrane proteins in wildtype *E. coli* cells.

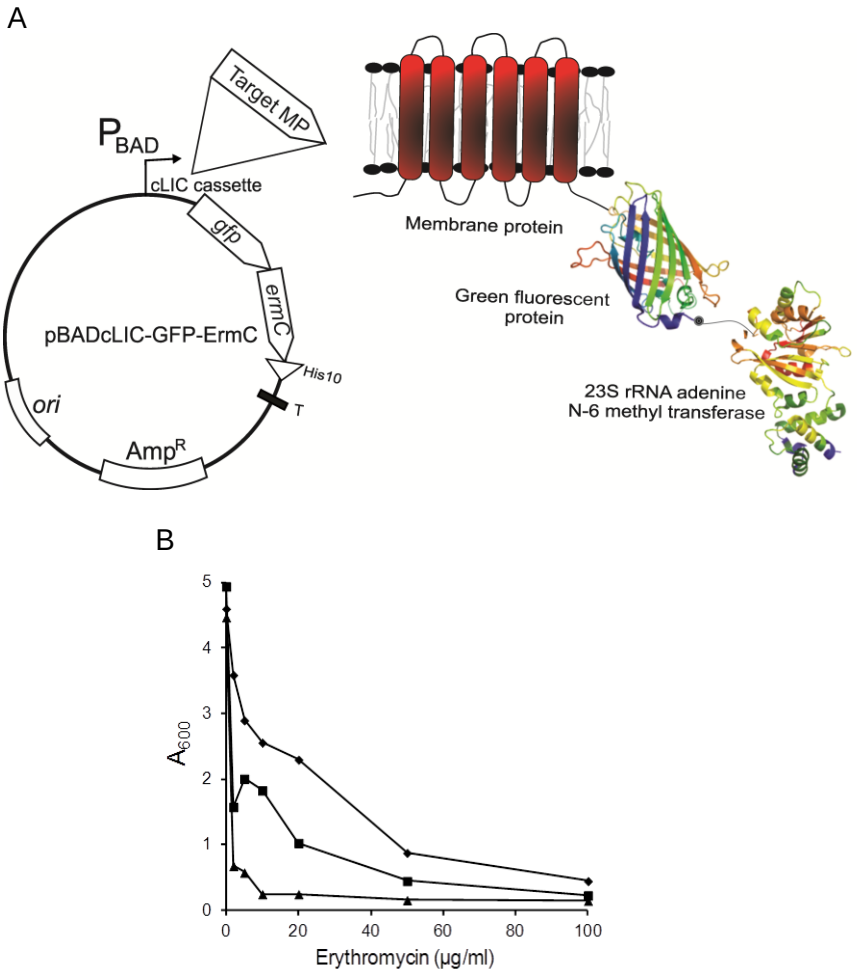

Figure 1: Selection system and strain used in the study.
(A) Plasmid-based selection system. The coding sequence of the target membrane protein is N-terminally fused in tandem to the folding reporter GFP and to a selectable marker encoding the 23S-rRNA adenine N-6-methyltransferase (ErmC). The expression cassette combines the P_{BAD} promoter with a high-throughput (ligation independent) cloning system.
(B) Sensitivity of *E. coli BW25113* (♦), *E. coli BW25113A* (■) *and E. coli BW25113B* (▲) to erythromycin (µg/ml), as determined from the A_{600} of liquid cultures after overnight growth in the presence of the indicated erythromycin concentrations.

3. Materials and methods

3.1.1 Strains and cultivation conditions for protein expression

The strains used in this study are listed in Supplementary Table 1. For cloning procedures *E. coli* MC1061 (Casadaban, Cohen 1980) was used as host, and *E. coli* BW25113 Δ*acrB ΔemrE mdfA*::kan (Tal, Schuldiner 2009) was used as host for the selection of strains with improved (evolved) expression performance. *E. coli* BW25113B is extremely sensitive to erythromycin because the genes encoding three major drug extrusion systems have been inactivated (Δ*acrB ΔemrE mdfA*::kan). *E. coli* strains carrying pBADcLIC derivatives were cultivated under vigorous aeration at 37 °C in Luria broth supplemented with 100 µg/ml ampicillin. In the case of *E. coli* BW25113 and derivatives, Luria broth was supplemented with 100 µg/ml ampicillin plus 5 µg/ml kanamycin. Cells were grown in Luria broth containing 100 µg/ml ampicillin *w/wo* 5 µg/ml kanamycin at 37°C in shake-flasks rotated at 200 rpm.

3.1.2 Selection and screening

E. coli BW25113B transformed with pBADc LacSΔIIA–GFP–ErmC, pBADc Gltp–GFP–ErmC and pBADcBcap–GFP–ErmC were grown in LB supplemented with 100 µg/ml ampicillin plus 5 µg/ml kanamycin. The temperature was lowered from 37°C to 25°C after the A_{600} had reached 0.5-0.6, and, after temperature equilibration, protein expression was induced by the addition of 0.01% (w/v) of L-arabinose. After 2 h of induction, a 2% (v/v) inoculum was introduced into fresh LB supplemented with 100 µg/ml ampicillin, 5 µg/ml kanamycin plus 5µg/ml erythromycin. Cultivation was continued for 48 h, after which the cells were diluted into fresh medium with an elevated erythromycin concentration (always using a 2% v/v culture inoculum). The erythromycin concentration was gradually increased from 5 to 10 to 20 to 50 to 100 to 200 µg/ml. After that cells were plated onto LB agar supplemented with 100 µg/ml ampicillin, 5 µg/ml kanamycin, 100 µg/ml erythromycin plus 0.01% w/v L-arabinose. Isolated colonies were selected for further analysis.

3.1.3 Construction of GFP-ErmC fusion proteins

DNA manipulations were done according to standard protocols. The vector pBADcLIC-ERY was obtained by insertion of the cLIC-ERY cassette from pREcLIC-ERY (Linares, Geertsma & Poolman 2010) into pBADcLIC (Geertsma, Poolman 2007) as *Nco*I-*Hin*dIII fragment. Plasmid pBADcLIC–GFP–ErmC was constructed in two steps. First, the GFP gene was amplified from pBADcLIC–GFP, using primer-introduced sites at the 5´ (*Swa*I) and 3´ (*Eco*RI-*Xba*I) ends; the gene was subsequently cloned as a *Swa*I-*Xba*I fragment in the pBADcLIC vector, yielding pBADcLIC–GFP*Eco*RI (the *Eco*RI site being created by the reverse primer). A *Eco*RI–XbaI fragment, corresponding to the *ermC*–His sequence,

amplified from pREcLIC–ErmC, was cloned into pBADc-LIC–GFP*Eco*RI, yielding pBADcLIC–GFP–ErmC.

3.1.4 Target proteins

The target genes coding for BcaP (branched-chain amino acid permease from *Lactococcus lactis*), LacSΔIIA (a C-terminal truncation mutant of the lactose transporter from *Streptococcus thermophilus*), PacL (putative P-type cation transporter from *L. lactis*), YidC (preprotein translocase from *L. lactis*), GerAC (spore germination factor from *Bacillus subtilis*), OxaA (YidC-like protein from *L. lactis*), GltP (glutamate transporter from *E. coli*) and Sav1866 (Multidrug ABC transporter from *Staphylococcus aureus*) were ligated into pBADcLIC–GFP–ErmC, using the ligation-independent cloning method (Geertsma, Poolman 2007). In all cases, the target protein was expressed with a C-terminal tobacco etch virus protease cleavage site (sequence ENLYFQG) followed by the GFP, ErmC and a 10 Histag.

3.1.5 Whole-cell fluorescence measurements

For whole-cell fluorescence measurements, cells expressing membrane protein-GFP-EmrC fusions were harvested, washed and subsequently resuspended to A_{600}=3 (1 mg/ml of total protein) in 100 mM KPi, pH 7. GFP emission was measured using a BioTek FL600 microplate fluorescence reader with the excitation wavelength set at 485 nm and the emission wavelength set at 530 nm (slit width of 20 and 25 nm, respectively). For direct comparison of the expression levels, the GFP fluorescence data were normalized to the same A_{600}. Background fluorescence levels were assessed by measuring whole-cell fluorescence of uninduced cells, and these values were subtracted.

3.1.6 In-gel fluorescence and anti-his tag immune-detection

Whole-cell samples corresponding to 1 mg of total protein were resuspended in 400 µl ice-cold 50 mM KPi, pH 7.2, 1 mM $MgSO_4$, 10% glycerol, 1 mM PMSF plus trace amounts of DNase. Glass beads (300 mg, 0.1 mm diameter) were added and samples were shaken in a Tissue Lyser LT (Qiagen) at 50 oscillations/min for 5 min. Aliquots (10 µl, 25 µg protein) were taken and 2.5 µl of 5× protein sample buffer [120 mM Tris–HCl, pH 6.8, 50% (v/v) glycerol, 100 mM DTT, 2% (w/v) SDS plus 0.1% (w/v) bromophenol blue] was added. Samples were stored on ice until use. Protein samples were analyzed by 10% SDS-PAGE electrophoresis and in-gel GFP fluorescence, using a Fujifilm LAS-3000 imaging system and AIDA software (Raytest; Isotopenmessgeräte, GmbH). Subsequently, the gels were submitted to semi-dry electroblotting and immunodetection with a primary antibody raised against a hexa-His tag (Amersham Pharmacia Biotech). Chemiluminescence detection was done using the Western-Light kit (Tropix Inc.) and quantified using the Fujifilm LAS-3000 imaging system.

3.1.7 Erythromycin resistance

The level of functional ErmC protein was estimated by quantification of bacterial survival as a function of erythromycin concentration. *E. coli* BW25113B cells expressing the different membrane protein-GFP-ErmC fusion were inoculated (2% v/v) into a fresh LB liquid medium with erythromycin concentrations varying from 0 to 100 µg/ml, and subsequently incubated overnight. The final optical density (A_{600}) reached by the culture was the parameter utilized to determine ErmC activity.

3.1.8 Plasmid copy number and plasmid curing

The plasmid copy number was determined by isolation of plasmid DNA, using the Promega miniprep kit, and estimation of the plasmid DNA concentration by Nanodrop read-out at 260 nm. For curing of evolved strains of pBADc Lic-MP-GFP-ErmC, the cells were grown at 42˚C in LB supplemented with 5 µg/ml kanamycin and subsequent transfer (0.1 % v/v inoculum) into pre-warmed media. After 25 rounds of sub-culturing, the cells were plated on LB agar plates and isolated colonies were further characterized by culturing on LB agar plates supplemented with and without 100 µg/ml ampicillin. Single colonies of cured clones were purified further and the cells were retransformed with the original plasmids and screened for expression performance.

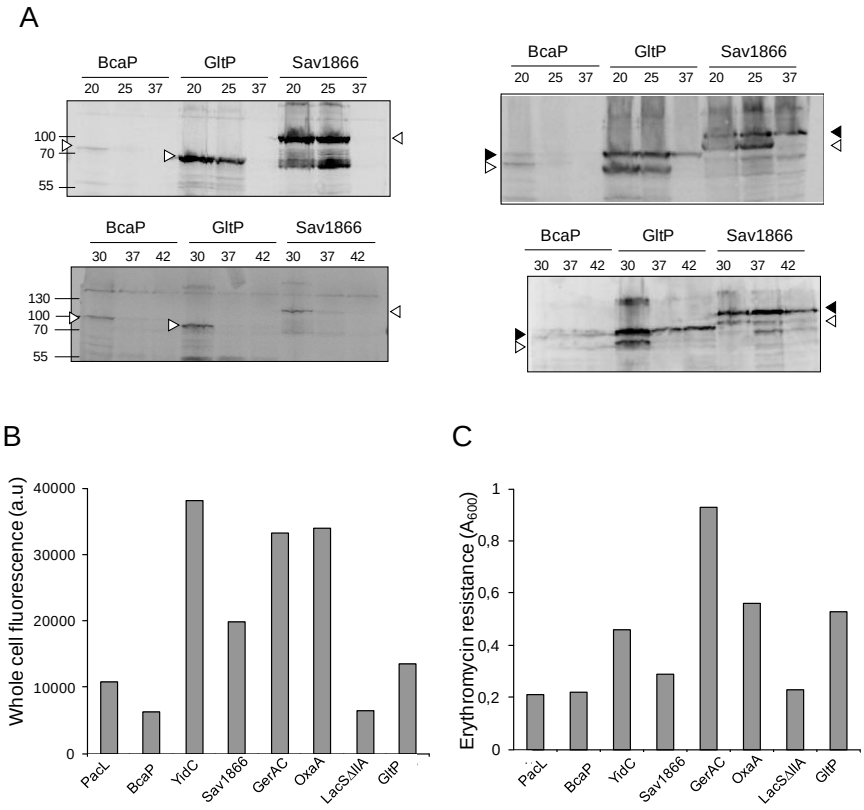

Figure 2:
(A) Optimal temperature for protein expression. Three membrane proteins, differing in basal level of expression in *E. coli* MC1061 were analyzed as GFP-ErmC fusions: BcaP (low level of expression), GltP (relatively high) and Sav1866 (medium). In-gel GFP fluorescence (*left panels*) and immunoblots probed with a anti-His tag antibody of the same gels (*right panels*) are shown. Black and white arrows indicate the positions of non-fluorescent (misfolded) and fluorescent (folded) protein species, respectively.
(B) *E. coli* BW25113B cells overexpressing membrane proteins as GFP-ErmC fusions. Cells were grown at 20 ºC on selective media with 50 µg/ml erythromycin and in the presence of 0.001% L-arabinose. No growth was observed in uninduced cultures.
(C) For each protein, the final OD_{600} of the culture at an erythromycin concentration of 50 µg/ml is given and an inducer concentration of 0.001% L-arabinose.

3.1.9 Genome sequencing and data analysis

Genomic DNA was purified using the Qiagen Genomic DNA Purification Kit. Shotgun DNA libraries were generated according to the manufacturer's sample preparation protocol for genomic DNA. Briefly, 1–5 μg of genomic DNA was randomly sheared using nebulizers and the ends were repaired using polynucleotide kinase and Klenow enzyme. The 5′ ends of the DNA fragments were phosphorylated and a single adenine base was added to the 3′ ends using Klenow exo+. Following ligation of a pair of Illumina adaptors to the repaired ends, the DNA was amplified in 10 cycles, using adaptor primers (Illumina, San Diego, CA), and fragments around 250 bp long were isolated from agarose gels. Sequencing libraries were quantified with a Bio-Rad Experion Analyzer as well as the Picogreen fluorescence assay. Cluster generations were performed on an Illumina cluster station using 4 pmol of sequencing libraries. Seventy-six cycles of sequencing were carried out using the Illumina GenomeAnalyzerII system according to the manufacturer's specifications. Sequencing analysis was first done using the Illumina analysis pipeline. The output of the Illumina analysis pipeline was fed into CLC Bio Software (Aarhus, Denmark). At the assembly stage, sequence reads of each mutant were aligned to the previously assembled wildtype (reference) genome.

4. Results

4.1 Choice of antibiotic resistance marker

The target (membrane) protein was fused to a C-terminal selectable marker that confers a drug resistance phenotype, when located in the cytoplasm. In the initial studies, we screened various antibiotic resistance markers, including the chloramphenicol (CAT), the tetracycline (TetA) and the kanamycin (Aph) resistance proteins as fusion partners for selection, but none of these markers proved useful in *E. coli*. Either the fusion partners were degraded or they dragged the target protein into a misfolded state. Because each of these antibiotic resistance proteins form an oligomeric complex (CAT and TetA are trimeric proteins and Aph is homodimeric), we reasoned that their quaternary structure and/or cellular location (TetA is a membrane protein) posed a problem, in particular when the membrane protein target would form an oligomer as well. We then identified the erythromycin-resistance protein (ErmC) as a monomeric, soluble antibiotic resistance protein (Linares, Geertsma & Poolman 2010). Fusions of membrane proteins with ErmC proved highly stable.

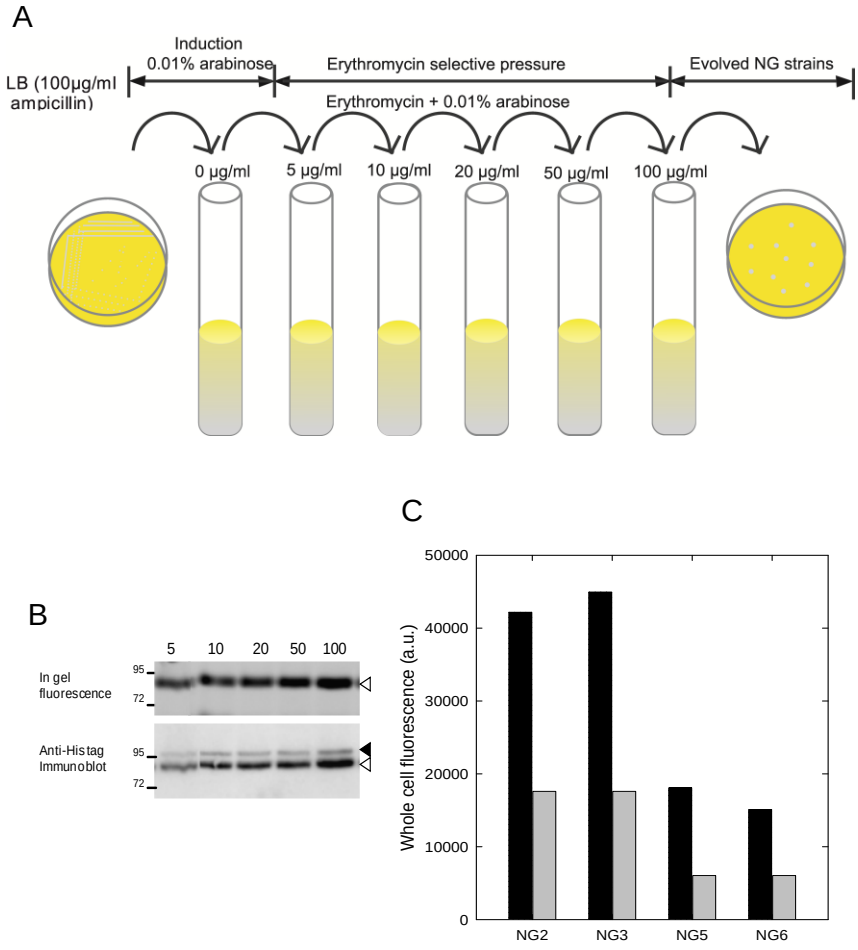

Figure 3:
(A) Scheme of the selection procedure.
(B) Selection of evolved strains. The in-gel GFP fluorescence and immunoblot data show the increase in LacSΔIIA expression with increasing erythromycin concentration. Samples were normalized based on OD_{600}; 25 µg total protein loaded/lane. Black and white arrows indicate the positions of non-fluorescent (misfolded) and fluorescent (folded) protein species, respectively.
(C) Comparison of best expressing clones (black) with parental strain BW25113B (grey); the NG strains were transformed with the same plasmid as used for evolving the strain: pBADc Lic-Gltp-GFP-ErmC in case of NG2 and NG3, pBADcLic-Bcap-GFP-ErmC in case of NG5 and NG6.

4.2 Choice of expression host

The first requirement for the suitability of ErmC-fusions is a high sensitivity of the expression host for erythromycin. As many *E. coli* strains are relatively insensitive to erythromycin, we evaluated isogenic strains with different capacity to excrete antibiotics, that are *E. coli* BW25113, *E. coli* BW25113A (Δ*acrB*) and BW25113B (Δ*acrB ΔemrE ΔmdfA*::kan) (Tal, Schuldiner 2009). The differences in growth in erythromycin-containing media were marked. *E. coli* BW25113 could grow up to 100 µg/ml of erythromycin, albeit that the final yield was reduced. By comparison, *E. coli* BW25113B was completely inhibited at 10 µg/ml erythromycin, whereas *E. coli* BW25113A displayed an intermediate phenotype (figure 1B). We selected *E. coli* BW25113B as expression host for the (membrane) protein-GFP-ErmC fusions, hereafter referred to as MP-GFP-ErmC.

4.3 The genetic selection system

Because ErmC can only be used as selection marker, we needed an additional protein partner to report functionality of the target protein. We thus constructed tandem fusions with GFP-ErmC C-terminal of the target protein. It has been established that GFP can serve as folding reporter and quantitatively assess the functionality of membrane proteins. In brief, when a protein, fused N-terminal to GFP, becomes misfolded during the biosynthesis, it turns GFP into a mis-folded, SDS-sensitive state. If the protein becomes properly folded, the GFP β-barrel will be assembled as a fluorescent SDS-resistant moiety; the SDS-sensitive and SDS-resistant conformations can be readily discriminated on SDS-PAGE and immunoblots. Whole-cell and in-gel fluorescence of protein extracts have been shown to match the activity of membrane proteins such as the lactose transporters from *Streptococcus thermophilus* (LacSΔIIA) and *E. coli* (LacY) and the glutamate transporter (GltP) from *E. coli* (Drew et al. 2006, Geertsma et al. 2008a).
MP-GFP-ErmC fusions were made with LacSΔIIA, BcaP, GltP, Sav1866, YidC and OxaA as target membrane protein. We observed two differently migrating protein species on immunoblots, the lower one being fluorescent in the in-gel assay (figure 2A). These data are consistent with previous observations on single fusions and show that a fraction of the target protein is well-folded and another fraction is misfolded. Thus, GFP can be used as reporter of the folding state of membrane proteins in dual fusions with ErmC. We also made fusions of GFP-ErmC with the water-soluble proteins GerAC (spore germination factor) and LmoSBP (substrate-binding protein).

4.4 Small-scale expression screening

To determine the optimal evolution screening conditions, we performed several small-scale expression studies. As variables we chose eight candidate membrane proteins, two *E. coli* hosts (MC1061 and BW25113), five temperatures (20, 25, 30,

37 and 42 ºC) and various inducer (L-arabinose) concentrations. The expression of the fusion proteins was evaluated by whole cell and in-gel fluorescence measurements and anti-His immunoblots. Figure 2A shows data from a representative set of experiments. For all the membrane proteins tested, the expression profiles were similar for the two *E. coli* strains (data not shown). The temperature at the time of production proved to be a key determinant for functional overexpression. The highest levels of functional expression (fluorescence) were generally associated with growth at relatively low temperature and either 20 or 25 ºC being optimal. At 37 and 42 ºC, hardly any fluorescence was observed and the fusion proteins were detected on immunoblots as a single band that is predicted to be misfolded.

A temperature 20 ºC in combination with the strain BW25113B and an inducer concentration of 0.001% L-arabinose was subsequently chosen as default condition to test the functionality of the MP-GFP-ErmC systems in a 96-well format, i.e. by comparing growth on selective media (with varying concentrations of erythromycin). In the presence of erythromycin without induction, none of the constructs supported growth. With 0.001% L-arabinose and in the presence of erythromycin, the cells expressing the target fusions survived. However, the final A_{600} reached varied, consistent with the different levels of expression of the proteins. As shown in figure 2B and 1C, cells overexpressing GerAC reached the highest cell denstity, *i.e.* A_{600} ~0.9 at 50 µg/ml of erythromycin. Cells expressing GltP, YidC and OxaA reached an intermediate A_{600} of ~0.5, whereas poorly expressed membrane proteins like BcaP, LacSΔIIA and PacL supported growth up to A_{600} of ~0.2. Thus, our selection system effectively discriminates between cells expressing high levels of target membrane protein and those expressing little or no protein.

4.5 Selection of improved expression hosts

For the selection of BW25113B derivatives with improved expression properties, we used LacSΔIIA, GltP and BcaP as target proteins and thus constructed pBADc LacSΔIIA–GFP–ErmC, pBADc GltP–GFP–ErmC and pBADcBcaP–GFP–ErmC. LacSΔIIA and BcaP express poorly and GltP expresses relatively well in BW25113B (figure 2B). *E. coli* BW25113B bearing a plasmid for expression of the different MP-GFP-ErmC fusions was grown in LB plus ampicillin and kanamycin, and the concentration of erythromycin was gradually increased. Cells were transferred to fresh medium every 48 h. Cultures resistant to 100 µg/ml erythromycin were plated on LB agar with 100 µg/ml erythromycin to obtain individual clones, which were then screened for folded membrane protein by monitoring the GFP fluorescence in whole cells (figure 3A and 3B) and analyzing selected samples on SDS-PAGE to confirm full-length expression of MP-GFP-ErmC fusions. Finally, four clones (NG2, NG3, NG5 and NG6) showing increased erythromycin resistance and increased GFP fluorescence were selected for a more complete analysis.

4.6 Plasmid characterization and curing

To exclude the possibility that the improved membrane protein expression relates to a mutation in the vector, *e.g.* causing altered plasmid copy, promoter activity or mutation in the target gene to be overexpressed, the plasmids were isolated from *E. coli* NG2, NG3, NG5 and NG6 and retransformed into *E. coli* BW25113B. The plasmid copy number was not altered and sequencing of the plasmids confirmed that the promoter region and genes were not mutated (figure 3C). Next, we cured *E. coli* NG2, NG3, NG5 and NG6 from its plasmids, and retransformed the evolved vector-free hosts with the parental plasmids. Testing of the expression levels of the fusion proteins in the evolved strains by whole cell and in-gel fluorescence and by immunoblotting confirmed that the alterations leading to enhanced protein production reside on the *E. coli* chromosome.

4.7 Establishing optimal expression conditions

To determine the optimal expression conditions, we performed several small-scale expression studies. First, the optimal inducer, L-arabinose, concentration was determined. Figure 4 shows the expression levels in the evolved and parental strains. NG3 and NG5 shows higher levels of expression compared to BW25113B but also the sensitivity for L-arabinose is shifted. In case of NG3 and NG5, 10^{-2} % (w/v) or higher concentrations of L-arabinose were required for maximal expression (figure 4). We observed that the expression levels increase with induction times. The highest levels of protein were generally found after overnight induction but in most cases high levels were already obtained after 5-6 h (figure 5).

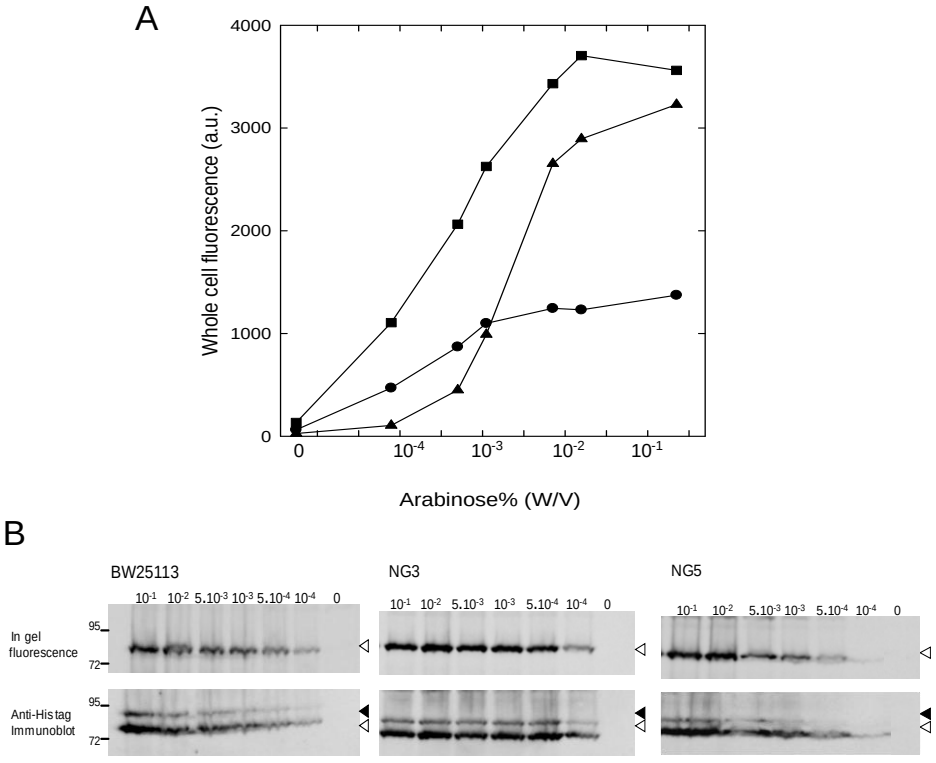

Figure 4: Screening for optimal inducer (L-arabinose) concentration of the expression of the glutamate transport protein (GltP). Two representative strains, NG3 (■) and NG5 (▲), are compared with the parental strain BW25113B (●). In-gel GFP fluorescence and immunoblots probed with an anti-His tag antibody of the same gels are shown. Black and white arrows indicate the positions of non-fluorescent (misfolded) and fluorescent (folded) protein species, respectively.

4.8 Protein expression in the evolved *E. coli* strains

The cured NG strains were re-transformed with pBADcBcaP–GFP–ErmC and pBADc GltP–GFP–ErmC to confirm that the acquired property for enhanced expression is located on the *E. coli* chromosome and to check for differences in expression patterns. The expression levels and global folding of the proteins were tested by whole cell fluorescence (total GFP fluorescence), in-gel fluorescence (to

discriminate between full length proteins and possible breakdown products), and anti-His tag immunoblots (to discriminate between folded and mis-folded fusion proteins (Geertsma et al. 2008a). Figure 6 shows that in each of the NG strains, the amount of fluorescent (folded) protein and the ratio of folded over mis-folded protein has increased relative to the parental strain BW25113B. Of the evolved strains, NG3 performed best overall, whereas NG2 did better in the overexpression GltP as compared to LacSΔIIA and BcaP. We note that NG2 was selected using GltP as target membrane protein, whereas NG5 and NG6 were obtained with BcaP. To test whether or not erythromycin is required for increased expression of the membrane protein-GFP-ErmC fusions, we performed expression trials with and without the antibiotic. Figure 6 shows that the expression of LacSΔIIA did not require erythromycin. Finally, we tested the expression of two soluble reporters, that are GerAC and LmoSBP, and we observed that the level of GerAC was the same in BW25113B and the NG strains, whereas the production of Lmo SBP was significantly increased in the evolved *E. coli* strains.

4.9 DNA sequencing of the evolved *E. coli* strains

Whole genome sequencing of the parent strain BW25113B and four evolved NG strains was done to explore the mechanism responsible for improved membrane protein production. The overall sequence coverage was more than 500-fold across the entire genome. The parent strain BW25113B was found to have 22 differences (mostly point mutations, 2 insertion mutations) relative to the sequence of *E. coli* K-12 sub-strain MG1655 (U00096.2). When compared to BW25113B, the evolved NG strains carried two or three mutations as summarized in Table 2. Remarkably, each of the NG strains had a mutation associated with *hns*, in the promoter region and/or the structural gene. The *hns* mutation in NG5 and NG6 leads to deletion of Thr25, Leu26, Glu27 and Glu28 in dimerization interface of HNS. This mutation is predicted to destabilize dimer formation. Because dimerization (or higher order oligomerization) must precede DNA-binding, the mutation will lead to a loss of function of H-NS. The evolved *E. coli* NG2 and NG3 strains have a nonsense mutation in *hns*. In addition, NG2 has a mutation in the promoter region of *hns*, which most likely arose prior to the nonsensense and only caused partial loss of H-NS function Strikingly and in accordance with the genotype of the *hns* mutations, NG2 and NG3 yielded the highest overexpression of the membrane proteins. It is possible that the secondary and/or tertiary mutations in the NG strains enhance or partially suppress the effects of the mutations in *hns* (Table 1 summarizes the characterization of four *E. coli* strains, evolved towards improved membrane protein production).

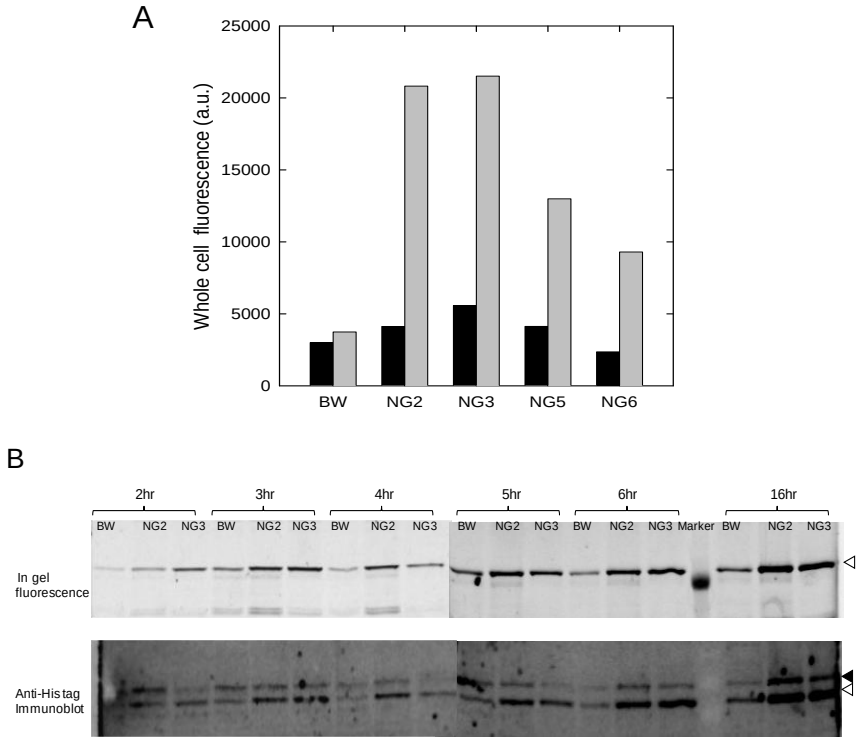

Figure 5: Optimization of expression conditions.
(A) Screening for optimal induction time at an L-arabinose concentration of 0.01% (w/v) of the expression of GltP. Samples were taken after 4 h (grey) and 16h (black) of induction.
(B) Comparison of expression of GltP in BW25113B, NG2 and NG3 at 0.01% (w/v) L-arabinose and different induction times. In-gel GFP fluorescence and immunoblots probed with an anti-His tag antibody of the same gels are shown. Black and white arrows indicate the positions of non-fluorescent (mis-folded) and fluorescent (folded) protein species, respectively.

Table 1: Mutations in evolved *E. coli* strains

Name	Genes	Coding region change	Amino acid change	Function of protein
NG2	*hns*	C376T	Gln126 to stop codon	Chromosome organization and transcriptional silencing
	hns promoter			Chromosome organization and transcriptional silencing
	ung	G220A	Gly74Ser	Uracil DNA glycosylase that modulates the mutation frequency; controlled by the *cpx* regulon
	cpxA	G524A	Ser175Asn	Histidine sensor kinase involved in cell envelope stress response
NG3	*hns*	C376T	Gln126 to stop codon	Chromosome organization and transcriptional silencing
	dcm	T1337A	Val1446Glu	DNA-cytosine methyltransferase
NG5	*hns*	Deletion from A73 to A84	Deletion of Thr25, Leu26, Glu27 and Glu28 in HNS dimerization domain	Chromosome organization and transcriptional silencing
	ydcU	C348A		Putative ABC transporter for spermidine and putrescine
	yffN	C195T		Unknown
NG6	*hns*	Deletion from A73 to A84	Deletion of Thr25, Leu26, Glu27 and Glu28 in HNS dimerization domain	Chromosome organization and transcriptional silencing
	YdcU	C348A		Putative ABC transporter for spermidine and putrescine

5. Discussion

The production of integral membrane proteins in a fully functional state still is a hit-or-miss affair. Ideally, one would like to tackle the expression problem without having to rely on optimization without trial-and-error. In this paper, we describe a simple and effective strategy to optimize the level of functional membrane protein. Our approach is based on a double translational fusion reporting system, using the *gfp* and *ermC* genes in tandem and fused 3' of the target gene. The GFP protein is used to monitor the folding state of the target protein; the ErmC protein (23S rRNA adenine N-6 methyl-transferase) is used to screen for increased expression by evolving strains towards increasing concentrations of erythromycin. A manageable screen for expression optimization can be performed in as little as 200 µl of culture and done for numerous target proteins in parallel. We previously focussed on *Lactococcus lactis* for enhanced membrane protein expression (Linares, Geertsma

& Poolman 2010). We now developed a strategy to evolve *E. coli* for enhanced production of integral membrane proteins, but the method is applicable for any type of protein. When the technology is used in its full potential, we foresee a panel of expression hosts, each evolved and optimized for a given class of proteins or tuned towards a specific protein.

Strategies for improved expression of membrane proteins, using selectable antibiotic resistance markers (chloramphenicol acetyltransferase (CAT) and kanamycin kinase (Aph) have been previously reported (Maxwell et al. 1999, Massey-Gendel et al. 2009). However, these methodologies do not allow the rapid quantification of the folded/misfolded protein fraction to be probed in parallel. The combination of the C-terminal ErmC tag with GFP provides a more advanced reporter system for protein expression. Moreover, the ErmC as fusion partner is superior over CAT and Aph and causes little or no deleterious effect on the target protein. A disadvantage of selecting for erythromycin resistance is that wildtype *E. coli* cells are already highly resistant to this antibiotic. We therefore used a triple drug/antibiotic export null strain (BW25113B) for the expression screening. By using whole cell and in-gel fluorescence measurements and immune-blot analysis, we determined the conditions for optimal expression in terms of the amount of fluorescent (functional) protein and the relative amounts of well-folded and mis-folded protein. We show that the fluorescence intensity of the target protein-GFP-ErmC fusion is a good indicator of the expression level. Importantly, the whole cell fluorescence correlates well with growth when analysed as a function of erythromycin concentration.

We obtained several evolved strains of which four were characterized in more detail, *i.e.* NG2, NG3, NG5 and NG6. NG3 proved most effective for all target proteins tested. In general, the evolved strains performed best for the target protein that was used in the initial screening. For instance, NG2 and NG3 yielded highest levels of expression of GltP, whereas NG5 and NG6 performed relatively best with BcaP. In contrast to the Walker strains (Miroux, Walker 1996), which yield increased amounts of biomass but little or no increase in the specific expression, the NG strains show increased expression per unit of biomass. Genome sequencing revealed that the NG strains have different mutations, but, remarkably, all have a mutation in *hns*. HNS is an abundant nucleoid-associated protein, involved in genome organization and transcriptional silencing of a variety of cellular functions (Bouvier et al. 1998, Dorman 2004, Zimmerman 2006). HNS consists of N- and C-terminal domains separated by a flexible linker region. The N-terminal part is composed of three α helices connected by linkers and this domain is required for dimerization of the protein (Bloch et al. 2003). The C-terminal domain of HNS binds to DNA and preferably to sites rich in AT (Dorman 2004); a TCGATAxATT consensus motif for high-affinity binding has been identified (Lang et al. 2007, Sette et al. 2009). HNS represses transcription by binding to high-affinity sites and cooperative spreading along DNA, thereby occupying the promoter region. As a consequence it also forms looped structures of DNA (Dame, Noom & Wuite 2006, Liu et al. 2010). In fact, HNS and other nucleoid associated

proteins (Fis, Dps and StpA) are thought to form large topological domain barriers, thereby determining the chromosome architecture. The interaction of HNS is affected by several environmental factors, incl. pH, osmolarity, temperature and growth phase, but differs from gene to gene and as such the regulation of gene expression is complex.

NG2 and NG3 have a nonsense mutation, which results in a defective HNS molecule. NG2 also has a mutation in the promoter region of *hns*. NG5 and NG6 have a mutation that results in the deletion of Thr25, Leu26, Glu27 and Glu28 in the N-terminal domain of HNS and most likely affects the dimerization of HNS. Since NG2 and NG3 as well as NG5 and NG6 have the same mutations in *hns*, it is likely that these arose first and that the secondary and tertiary mutations were formed later. NG5 and NG6 also have the same secondary mutation (in *ydcU*), and it seems likely that NG5 is derived from NG6. Below, we briefly discuss the functions of the genes having a secondary and/or tertiary mutation in the NG strains.

NG2 has substitution mutation in *ung and cpAx* in addition to a nonsense mutation in *hns.* CpxA is the histidine sensor kinase of a two-component signal transduction system. The Cpx two-component regulatory system responds to pH and osmotic stress, as well as aggregated and misfolded proteins (Pogliano et al. 1997). Cpx activates the expression of protein folding catalysts and protein degradation factors (Duguay, Silhavy 2004), and given its role in the cell envelop stress response it could well be an important factor for functional expression of membrane proteins. Phosphorylated CpxR, the response regulator of the Cpx system, negatively regulates *ung* transcription (Ogasawara et al. 2004). *ung* codes for uracil DNA glycosylase, which removes uracil from DNA via base excision repair and thus diminishes the mutation level of *E. coli*.

NG3 has an additional mutation in *dcm*, which codes for DNA-cytosine methyltransferase, an enzyme involved in DNA methylation. Methylated cytosine is subject to deamination, which causes C to T mutations. This mismatch is repaired by the very short patch (VSP) mismatch repair pathway. Vsr is an essential component of VSP and its activity counterbalances the mutagenesis associated with *dcm* activity (Macintyre, Doiron & Cupples 1997). The chromosomal arrangement of *dcm* and *vsr* is unusual and the 5' end of *vsr* overlaps the 3' end of *dcm* in 1 reading frame (Sohail et al. 1990). The two genes are probably transcribed as a single mRNA, with translation of *vsr* depending on translation of *dcm* (Dar, Bhagwat 1993). This gene arrangement may assure that Vsr is produced along with Dcm and may minimize the mutagenic effects of cytosine methylation. NG3 has a mutation in the 3' region of *dcm*, which might affect VSP-mediated mismatch repair.

NG5 and NG6 have the same mutation in *ydcU* in addition to a mutation that affects the dimerization of HNS. YdcU is the membrane component of a binding-protein-dependent ABC transport system that takes up spermidine/putrescine, but it is unclear how it could have impact on the functional expression of membrane proteins. Comparison of NG5 and NG6 with an isogenic strain only carrying the

hns mutation will indicate to what extent *ydcU* plays a role in the overexpression of functional membrane protein.

Table 2: *E. coli* strains used in this study

Strains	Genotype	Reference
MC1061	*ara*D139, *Δ(ara, leu)* 7697, *Δlac*X 74, *gal*U^{-}, galK^{-}, *hsr*$^{-}$, *hsm*$^{+}$, *str*A	(Casadaban, Cohen 1980)
BW25113	*rrnB3 ΔlacZ4787 hsdR514* Δ(*araBAD*)567 Δ*(rhaBAD)*568 *rph-1*	(Datsenko, Wanner 2000)
BW25113A	BW25113 Δ*acrB*::kan	(Tal, Schuldiner 2009)
BW25113B	BW25113 Δ*acrB* Δ*emrE* Δ*mdfA*::kan	(Tal, Schuldiner 2009)
NG2	BW25113B *hns1, hns2, ung1, cpxA*	This study
NG3	BW25113B *hns1, dcm1*	This study
NG5	BW25113B *hns3, ydcU1, yffN1*	This study
NG6	BW25113B *hns3, ydcU1*	This study

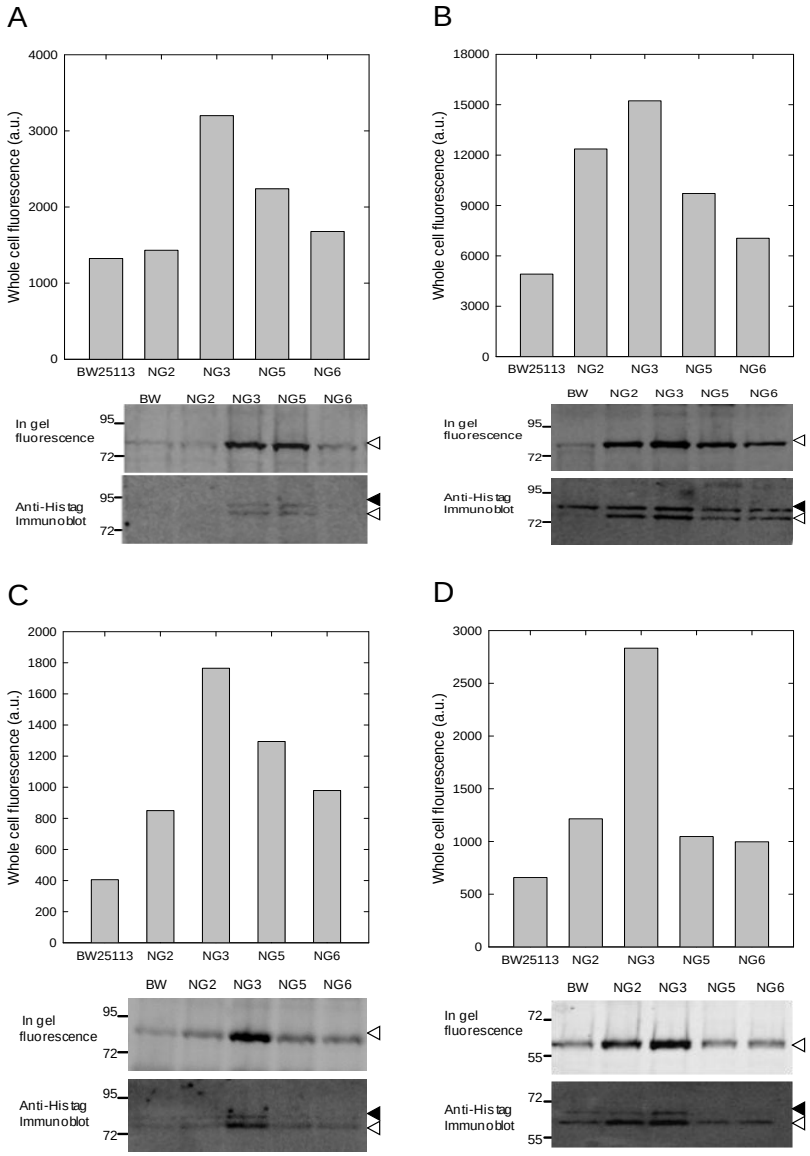

Figure 6: Three representative membrane proteins expressed in the NG strains. L-arabinose at 0.01% (w/v) was used for overnight induction. In-gel GFP fluorescence and immunoblots probed with an anti-His tag antibody of the same gels are shown. Black and white arrows indicate the positions of non-fluorescent (mis-folded) and fluorescent (folded) protein species, respectively.
(A) BcaP-GFP-ErmC, **(B)** GltP-GFP-ErmC, **(C)** LacSΔIIA-GFP-ErmC **(D)** LacSΔIIA-GFP.

6. Conclusion

The strategy presented here allows rapid screening of amplified expression of almost any (membrane) protein, which will generate an even larger variety of better expressing hosts. Combined with the ease nowadays to sequence prokaryotic genomes, it will generate a wealth of information on what limits the functional expression of membrane proteins. It will eventually lead to rules regarding expression bottlenecks for different classes of integral membrane proteins. The strains isolated and characterized will be made available for use by the scientific community.